1. 行列转换

1.1. 需求分析

// 建表语句：

// 字段解释：id,学号，课程，成绩

1.2. MySQL处理

```sql
-- MySQL建表
CREATE TABLE course(
  id int(11) NOT NULL AUTO_INCREMENT PRIMARY KEY,
  sid int(11) DEFAULT NULL,
  course varchar(255) DEFAULT NULL,
  score int(11) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

// 插入数据

```sql
INSERT INTO course VALUES (1, 1, 'yuwen',  43);
INSERT INTO course VALUES (2, 1, 'shuxue', 55);
INSERT INTO course VALUES (3, 2, 'yuwen',  77);
INSERT INTO course VALUES (4, 2, 'shuxue', 88);
INSERT INTO course VALUES (5, 3, 'yuwen',  98);
INSERT INTO course VALUES (6, 3, 'shuxue', 65);
INSERT INTO course VALUES (7, 3, 'yingyu', 80);
```

1.3. Hive处理

```sql
-- Hive建表
create database if not exists exercise_db;
use exercise_db;
drop table if exists exercise_course;
create table exercise_course(id int, sid int, course string, score int) row
format delimited fields terminated by ',';
load data local inpath "/home/bigdata/exercise_course.txt" into table
exercise_course;
select * from exercise_course;
```

Hive表数据文件：exercise_course.txt

```
1,1,yuwen,43
2,1,shuxue,55
3,2,yuwen,77
4,2,shuxue,88
5,3,yuwen,98
6,3,shuxue,65
7,3,yingyu,80
```

求：所有数学课程成绩 大于 语文课程成绩的学生的学号

```
sid,yuwen,shuxue
1,43,55
2,77,88
3,98,65,80
```

1.4. 解决方案

原始数据格式：

```
select id,sid,course,score from exercise_course;
```

结果：

```
+-----+------+--------+-------+
| id  | sid  | course | score |
+-----+------+--------+-------+
| 1   | 1    | yuwen  | 43    |
| 2   | 1    | shuxue | 55    |
| 3   | 2    | yuwen  | 77    |
| 4   | 2    | shuxue | 88    |
| 5   | 3    | yuwen  | 98    |
| 6   | 3    | shuxue | 65    |
| 7   | 3    | yingyu | 80    |
+-----+------+--------+-------+
```

解题：

第一步：行列转换：

```
select sid,
max(case course when "yuwen" then score else 0 end) as yuwen,
max(case course when "shuxue" then score else 0 end) as shuxue,
max(case course when "yingyu" then score else 0 end) as yingyu
from exercise_course group by sid;
```

结果：

```
+------+--------+---------+---------+
| sid  | yuwen  | shuxue  | yingyu  |
+------+--------+---------+---------+
| 3    | 98     | 65      | 80      |
| 1    | 43     | 55      | 0       |
| 2    | 77     | 88      | 0       |
+------+--------+---------+---------+
```

第二步：结果SQL求得最终结果

```sql
select aa.sid from (
select sid,
max(case course when "yuwen" then score else 0 end) as yuwen,
max(case course when "shuxue" then score else 0 end) as shuxue,
max(case course when "yingyu" then score else 0 end) as yingyu
from exercise_course group by sid
) aa where aa.shuxue > aa.yuwen;
```

第三步：结果：

```
+---------+
| aa.sid  |
+---------+
| 1       |
| 2       |
+---------+
```